

# Web-based Supplemental Materials for: “Simultaneous Inference of Treatment Effect Modification by Intermediate Response Endpoint Principal Strata with Application to Vaccine Trials”

Yingying Zhuang, Ying Huang, Peter B. Gilbert

## A The estimating function for $\gamma$

The likelihood function for  $P(W|S)$  can be derived as

$$L(\gamma) = \prod_{i=1}^N p_{1i}^{g_{1i}} \times p_{2i}^{g_{2i}} \times \cdots \times p_{Di}^{g_{Di}}.$$

Taking the log and using the fact that  $\sum_j g_{ji} = 1$  for each  $i$ , the log-likelihood function is

$$l(\gamma) = \sum_{i=1}^n [y_{1i} \ln p_{1i} + \cdots + y_{D-1i} \ln p_{D-1i} - \ln(1 + p_{1i} + \cdots + p_{D-1i})].$$

The likelihood equations are found by taking the first partial derivatives of  $l(\gamma)$  with respect to each of the  $(D-1) \times (T+1)$  unknown parameters. The general form of these equations is:  $\frac{\partial l(\gamma)}{\partial \gamma_{jt}} = \sum_{i=1}^N h_t(s_{1i})(g_{ji} - p_{ji})$  for  $j = 1, 2, \dots, D-1$  and  $t = 0, 1, 2, \dots, T$ .

## B Asymptotic Distribution for the proposed VE estimator $\widehat{\text{VE}}^{(new)}(s_1)$

### B.1 Regularity conditions to be satisfied:

1.  $\text{risk}_{(z)}(S, W) > 0$  and  $P(\delta = 1|S, W) > 0$  almost surely.
2.  $\text{risk}_{(z)}(S, W; \beta)$  is thrice differentiable with respect to  $\beta$ . For  $\beta$  in a neighborhood of the true value  $\beta_0$ , the third derivatives are bounded by an integrable function of  $(Y, Z, S, W)$ .

3.  $P(\delta = 1|Y, Z, W)$  as a function of  $\alpha$  is thrice differentiable with respect to  $\alpha$ . For  $\alpha$  in a neighborhood of the true value  $\alpha_0$ , the third derivatives are bounded by an integrable function of  $(Y, Z, S, W)$ .

4.  $\Psi_\beta$  is nonsingular.

5. For all  $y, w, z$ , under  $\beta_0$  and  $\alpha_0$ ,

$$0 < \int \frac{P(y|s, z, w)}{P(\delta = 1|s, w)} dF(s|w, \delta = 1) < \infty$$

$$0 < \int |U_\beta(y|s, z, w)| \frac{P(y|s, z, w)}{P(\delta = 1|s, w)} dF(s|w, \delta = 1) < \infty.$$

6. For all  $y, s, w, z$ ,  $P(y|s, z, w)/P(\delta = 1|s, w)$  and  $U_\beta(y|s, z, w)P(y|s, z, w)/P(\delta = 1|s, w)$  are twice differentiable with respect to  $\beta, \alpha$ , with the second derivatives uniformly integrable with respect to  $F(s|w, \delta = 1)$  for  $(\beta, \alpha)$  within a neighborhood of  $(\beta_0, \alpha_0)$ .

7.  $\Psi_\gamma$  is nonsingular.

8.  $P \|\Psi_{\gamma_0}\|^2 < \infty$  and that the map  $\gamma \mapsto P\Psi_\gamma$  is differentiable at a zero  $\gamma_0$ , with a nonsingular derivative matrix.

For convenient notation, we let  $\pi_0 = \pi_{\alpha_0}(Y, Z, W) = P(\delta = 0|Y, Z, W; \alpha_0)$ , and  $\hat{\pi} = \pi_{\hat{\alpha}}(Y, Z, W) = P(\delta = 0|Y, Z, W; \hat{\alpha})$ .

9.  $P \|\Psi_{\alpha_0}\|^2 < \infty$  and that the map  $\alpha \mapsto P\Psi_\alpha$  is differentiable at a zero  $\alpha_0$ , with a nonsingular derivative matrix.

10. For  $\alpha$  in a neighborhood of  $\alpha_0$ , where  $\zeta > 0$  and  $\psi$  satisfies  $E\psi^2 < \infty$ :

$$\left| \frac{1}{\hat{\pi}} - \frac{1}{\pi_0} - \frac{-\dot{\pi}_0^T}{\pi_0^2}(\alpha - \alpha_0) \right| \leq \psi |\alpha - \alpha_0|^{1+\zeta}.$$

Conditions 1–6 strictly followed HGW for the asymptotic distribution of the pseudo-score estimator  $\hat{\beta}$ . Conditions 7–8 are needed to establish the asymptotic distribution

of the WL estimator  $\hat{\gamma}$ . Conditions 9–10 are needed for estimation with estimated sampling weights, where we substitute  $\pi_0$  with a consistent estimator  $\hat{\pi}_\alpha$ , where  $\alpha$  is a parameter to be estimated by ML from the Phase-I observations. Condition 10 typically follows from Condition 9 provided that  $\pi_\alpha$  has a continuous second derivative.

## B.2 Asymptotic normality:

Theorem 1: Under specified regularity conditions, as the sample size  $N \rightarrow \infty$ , we have

1.

$$\sqrt{N}(\hat{\beta} - \beta) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_1(\delta_i, Y_i, S_i, Z_i, W_i) + o_p(1) \rightarrow_d N(0, V), \quad (1)$$

2.

$$\sqrt{N}(\hat{\gamma}(\hat{\alpha}) - \gamma) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_2(\delta_i, Y_i, S_i, Z_i, W_i) + o_p(1) \rightarrow_d N(0, K), \quad (2)$$

3.

$$\sqrt{N} \left( \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} - \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \right) \rightarrow_d N \left( \mathbf{0}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right), \quad (3)$$

where  $\Sigma_{11} = V, \Sigma_{22} = K, \Sigma_{12} = \Sigma_{21} = \text{cov}(\phi_1(\delta, Y, S, Z, W), \phi_2(\delta, Y, S, Z, W))$ .

Sketch of Proof:

### B.2.1 Asymptotic normality of $\hat{\beta}$ :

The pseudo-score estimator  $\hat{\beta}$  is then obtained by solving the equation  $U(\beta, F_N, \hat{\pi}) = 0$ , and HGW proved that

$$\begin{aligned} \sqrt{N}(\hat{\beta} - \beta) &= -\dot{\Psi}_\beta^{-1}(\beta_0, F_0^*, \pi_0) \sqrt{N} \left\{ \Psi_N(\beta_0; F_0^*, \pi_0) + \sum_{i=1}^K \Psi_{F_k^*}[F_{Nk} - F_{k0}^*] + \Psi[\hat{\alpha} - \alpha_0] \right\} + o_p(1) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_1(\delta_i, Y_i, S_i, Z_i, W_i) + o_p(1), \end{aligned}$$

where

$$\phi_1(\delta, Y, S, Z, W) = a_0(\delta, Y, S, Z, W) + a_1(\delta, S, W) + a_2(\delta, Y, Z, W),$$

with

$$a_0(\delta, Y, S, Z, W) = \delta U_\beta(Y|S, Z, W) + (1 - \delta)E \{U_\beta(Y|S, Z, W)|Y, Z, W, \delta = 1\},$$

$$a_1(\delta, S, W) = \sqrt{N} \dot{\Psi}_{F_k^*}(\beta_0, F_0^*)(F_{Nk}^* - F_{k0}^*),$$

$$a_2(\delta, Y, Z, W) = \dot{\Psi}_\alpha(\beta_0, F_0^*, \pi(\alpha_0))I_\alpha^{-1}U_\alpha(\delta|Y, Z, W) + o_p(1),$$

and

$$V = \dot{\Psi}_\beta^{-1}Var(\phi_0)\dot{\Psi}_\beta^{-t}.$$

### B.2.2 Asymptotic normality of $\hat{\gamma}$ :

The weighted likelihood estimator  $\hat{\gamma}$  is obtained by solving:

$$\Psi_N(\gamma) \equiv \frac{\partial l(\gamma)}{\partial \gamma_{jt}} = \frac{1}{N} \sum_{i=1}^N \frac{\delta_i}{\pi_0(y_i, z_i, w_i)} h_t(s_{1i})(g_{ji} - p_{ji}) = 0.$$

Furthermore,

$$\frac{\partial^2 l(\gamma)}{\partial \gamma_{jt} \partial \gamma_{jt'}} = - \sum_{i=1}^N \frac{\delta_i}{\pi_0(y_i, z_i, w_i)} h_{t'}(s_{1i}) \cdot h_t(s_{1i}) \cdot p_{ji} \cdot (1 - p_{ji}) \quad (4)$$

$$\frac{\partial^2 l(\gamma)}{\partial \gamma_{jt} \partial \gamma_{j't'}} = \sum_{i=1}^N \frac{\delta_i}{\pi_0(y_i, z_i, w_i)} h_{t'}(s_{1i}) \cdot h_t(s_{1i}) \cdot p_{ji} \cdot p_{j'i}. \quad (5)$$

The information matrix  $I(\gamma_0)$  can be estimated by the observed information matrix  $\hat{I}(\hat{\gamma})$ , whose elements are the negatives of the values in equations (4) and (5) evaluated at  $\hat{\gamma}$ .

Now following similar steps in (Breslow & Wellner 2006), we apply Theorem 19.26 of van der Vaart (1998) to conclude that

$$\sqrt{N}(\hat{\gamma} - \gamma_0) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\delta_i}{\pi_0(y_i, z_i, w_i)} \tilde{l}_0(s_{1i}) + o_p(1)$$

where  $\tilde{l}_0$  denotes the efficient influence function  $\tilde{l}_0(s_{1i}) = I^{-1}(\gamma_0)\dot{l}_0(\gamma_0|s_{1i})$ .

The asymptotic variance is therefore:

$$\begin{aligned}
Var\sqrt{N}(\hat{\gamma} - \gamma_0) &= Var\left(\frac{\delta}{\pi_o}\tilde{l}_0\right) \\
&= VarE\left(\frac{\delta}{\pi_o}\tilde{l}_0|S(1)\right) + EVar\left(\frac{\delta}{\pi_o}\tilde{l}_0|S(1)\right) \\
&= Var(\tilde{l}_0) + E\left[\frac{\tilde{l}_0^{\otimes 2}}{\pi_o^2}Var(\delta|S(1))\right] \\
&= I(\gamma_0)^{-1} + P_0\left(\frac{1 - \pi_0}{\pi_0}\tilde{l}_0^{\otimes 2}\right).
\end{aligned}$$

**Estimation with estimated sampling weights:** In this section, we show that under mild assumptions,

$$\begin{aligned}
\sqrt{N}[\hat{\gamma}(\hat{\alpha}) - \gamma_0] &= \sqrt{N}[\hat{\gamma}(\hat{\alpha}) - \hat{\gamma}(\alpha_0)] + \sqrt{N}[\hat{\gamma}(\alpha_0) - \gamma_0] \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_2(\delta_i, Y_i, S_i(1), Z_i, W_i) + o_p(1) \rightarrow_d N(0, K). \quad (6)
\end{aligned}$$

When we substitute  $\pi_0$  with a consistent estimator  $\hat{\pi}_\alpha$ , where  $\alpha$  is a parameter to be estimated by ML from the Phase-I observations, under regularity Condition 9, the ML estimator  $\hat{\alpha}$  is consistent and asymptotically normal with influence function  $\tilde{l}_0^\alpha$  so that

$$\sqrt{N} \begin{pmatrix} \hat{\gamma}(\alpha_0) - \gamma_0 \\ \hat{\alpha} - \alpha_0 \end{pmatrix} = \sqrt{N} \begin{pmatrix} \frac{1}{N} \sum_{i=1}^N \frac{\delta_i}{\pi_{0i}} \tilde{l}_0(s_{1i}) \\ \frac{1}{N} \sum_{i=1}^N \tilde{l}_0^\alpha \end{pmatrix}. \quad (7)$$

Moreover,

$$\begin{aligned}
&\frac{1}{N} \sum_{i=1}^N \left(\frac{\delta_i}{\hat{\pi}_i} - \frac{\delta_i}{\pi_0}\right) \tilde{l}_0(s_{1i}) \\
&= \frac{1}{N} \sum_{i=1}^N \delta_i \tilde{l}_0(S_i) \left[ \frac{1}{\hat{\pi}_i} - \frac{1}{\pi_0} - \frac{-\pi_0^T}{\pi^2}(\hat{\alpha} - \alpha_0) \right] + \frac{1}{N} \sum_{i=1}^N \delta_i \tilde{l}_0(s_{1i}) \left[ \frac{-\pi_0^T}{\pi^2}(\hat{\alpha} - \alpha_0) \right] \\
&= R_N - \frac{1}{N} \sum_{i=1}^N \delta_i \tilde{l}_0(s_{1i}) \left[ \frac{\pi_0^T}{\pi^2}(\hat{\alpha} - \alpha_0) \right]. \quad (8)
\end{aligned}$$

By regularity Condition 10,

$$\begin{aligned}
|R_N| &= \frac{1}{N} \sum_{i=1}^N \delta_i \tilde{l}_0(S_i) \left[ \frac{1}{\hat{\pi}_i} - \frac{1}{\pi_0} - \frac{-\dot{\pi}_0^T}{\pi^2} (\hat{\alpha} - \alpha_0) \right] \\
&\leq \frac{1}{N} \sum_{i=1}^N \psi \left| \tilde{l}_0(S_i) \right| |\hat{\alpha} - \alpha_0|^{1+\zeta} \\
&= O_p(1) |\hat{\alpha} - \alpha_0| |\hat{\alpha} - \alpha_0|^\zeta = O_p(1) O_p(N^{-1/2}) o_p(1).
\end{aligned}$$

Multiplying through (8) by  $\sqrt{N}$ , we conclude that equation (6) holds by virtue of  $\sqrt{N}R_N = o_p(1)$  and the strong law of large numbers.

Furthermore,

$$K = Var \sqrt{N} [\hat{\gamma}(\hat{\alpha}) - \gamma_0] = Var \left( \frac{\delta}{\pi_0} \tilde{l}_0 - \frac{\delta}{\pi_0} \cdot \frac{\tilde{l}_0 \dot{\pi}_0^T}{\pi_0} \left( \frac{\delta}{\pi_0} \cdot \frac{\dot{\pi}_0^{\otimes 2}}{\pi_0(1 - \pi_0)} \right)^{-1} \frac{\delta}{\pi_0} \cdot \frac{\dot{\pi}_0 \tilde{l}_0^T}{\pi_0} \right).$$

### B.2.3 Asymptotic normality of $(\hat{\beta}, \hat{\gamma})$ :

$$\sqrt{n} \left( \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} - \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \right) \xrightarrow{d} N \left( \mathbf{0}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \quad (9)$$

where  $\Sigma_{11} = V, \Sigma_{22} = K, \Sigma_{12} = \Sigma_{21} = cov(\phi_1(\delta, Y, S, Z, W), \phi_2(\delta, Y, S, Z, W))$ .

The proposed VE estimator  $\widehat{VE}^{(new)}(s_1)$  is a continuous function of  $\hat{\beta}$  and  $\hat{\gamma}$ ; therefore, the regular delta method applies.

## C Theoretical justification for perturbation resampling methods

### C.1

In this section, we show that

$$\begin{aligned}
\sqrt{N}(\hat{\beta}^{(\epsilon)} - \beta) &\equiv -[\dot{\Psi}_\beta^{(\epsilon)}]^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_1(\delta_i, Y_i, S_i, Z_i, W_i)^{(\epsilon)} + o_p(1) \\
&= -[\dot{\Psi}_\beta]^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_1(\delta_i, Y_i, S_i, Z_i, W_i) \epsilon_i + o_p(1). \quad (10)
\end{aligned}$$

We shall use  $\mathcal{F}$  to denote the  $\sigma$ -field generated by the original data  $(\delta_i, Y_i, S_i, Z_i, W_i)$ . First, consider the unconditional version of  $\hat{\beta}^{(\epsilon)}$  with respect to the joint probability space of  $\mathcal{F}$  and  $\epsilon_i (i = 1, \dots, N)$ :

Under suitable equicontinuity conditions and smoothness conditions (van der Vaart and Wellner, 1996), we have:

$$\begin{aligned}\hat{W}_s^{(\epsilon)} &= \sqrt{N}(\hat{\beta}^{(\epsilon)} - \beta_0) \\ &= \left[ -\dot{\Psi}_\beta^{(\epsilon)}(\beta_0, F_0^*, \pi_0) \right]^{-1} \sqrt{N} \left\{ \Psi_N^{(\epsilon)}(\beta_0; F_0^*, \pi_0) + \sum_{i=1}^K \Psi_{F_k^*}^{(\epsilon)}[F_{Nk}^{(\epsilon)} - F_{k0}^*] + \dot{\Psi}_\alpha^{(\epsilon)}[\hat{\alpha} - \alpha_0] \right\} + o_p(1) \\ &= \left[ -\dot{\Psi}_\beta^{(\epsilon)} \right]^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_1^{(\epsilon)}(\delta_i, Y_i, S_i, Z_i, W_i) + o_p(1),\end{aligned}$$

where  $\phi_1^{(\epsilon)} = a_0^{(\epsilon)} + a_1^{(\epsilon)} + a_2^{(\epsilon)}$ .

To show that  $\hat{W}_s^{(\epsilon)} = \sqrt{N}(\hat{\beta}^{(\epsilon)} - \beta_0) = -[\dot{\Psi}_\beta]^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_1(\delta_i, Y_i, S_i, Z_i, W_i) \epsilon_i + o_p(1)$ , it suffices to show the following:

1.  $\dot{\Psi}_\beta^{(\epsilon)} = \dot{\Psi}_\beta$
2.  $a_0^{(\epsilon)} = \epsilon a_0$
3.  $a_1^{(\epsilon)} = \epsilon a_1 + o_p(1)$
4.  $a_2^{(\epsilon)} = \epsilon a_2 + o_p(1)$ .

**1. To show that  $\dot{\Psi}_\beta^{(\epsilon)} = \dot{\Psi}_\beta$**

$$\begin{aligned}\dot{\Psi}_\beta^{(\epsilon)}(\beta_0, F_0^*, \pi_0) &= \frac{\partial}{\partial \beta} \Psi^{(\epsilon)}(\beta_0, F_0^*, \pi_0) \\ &= -E_0[\epsilon \delta I_\beta(Y|S, Z, W) - \epsilon(1 - \delta) \frac{\partial \int U_\beta(Y|s, Z, W) h(s|Y, Z, W) ds}{\partial \beta}].\end{aligned}$$

Since  $\epsilon$  is independent of  $(\delta, Y, S, Z, W)$  and  $E(\epsilon) = 1$ ,  $\dot{\Psi}_\beta(\beta_0, F_0^*, \pi_0) = \dot{\Psi}_\beta^{(\epsilon)}(\beta_0, F_0^*, \pi_0)$ .

**2. To show that**  $a_0^{(\epsilon)} = \epsilon a_0$

$a_0(\delta, Y, S, Z, W)^{(\epsilon)} = a_0(\delta, Y, S, Z, W)$  by definition of  $\hat{\beta}^{(\epsilon)}$ .

**3. To show that**  $a_1^{(\epsilon)} = \epsilon a_1 + o_p(1)$

We first derive the general asymptotic properties of a perturbed MLE estimator:

$L(\beta) = \prod_{i=1}^n p_\beta(x_i)$ ,  $l(\beta) = \sum_{i=1}^n \log p_\beta(x_i)$ ,  $\frac{\partial}{\partial \beta} l(\beta) \equiv \dot{l}(\beta) \equiv U(\beta)$ , and the MLE  $\hat{\beta}$  is obtained by solving the estimating equation  $\sum_{i=1}^n U(\beta|X_i) = 0$ ,

$$0 = \frac{1}{\sqrt{n}} \dot{l}_n(\hat{\beta}) = \frac{1}{\sqrt{n}} \dot{l}_n(\beta_0) - \left(-\frac{\ddot{l}_n(\beta_n^*)}{n}\right) \sqrt{n}(\hat{\beta} - \beta_0), \text{ where } |\beta_n^* - \beta_0| \leq |\hat{\beta} - \beta_0|. \text{ And } -\frac{1}{n} \ddot{l}_n(\beta_n^*) = -\frac{1}{n} \ddot{l}_n(\beta_0) + o_p(1) \xrightarrow{p} I(\beta_0).$$

$$\text{Thus } \sqrt{n}(\hat{\beta} - \beta_0) = I^{-1}(\beta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_n(\beta_0|X_i) + o_p(1) \xrightarrow{d} N(0, I^{-1}(\beta_0)).$$

Now define the perturbed MLE  $\hat{\beta}^{(\epsilon)}$  as the solution of

$$\sum_{i=1}^n \dot{l}^{(\epsilon)}(\beta|X_i) = 0, \text{ where } \dot{l}^{(\epsilon)}(\beta|X_i) = \epsilon_i \dot{l}(\beta|X_i).$$

$$\text{Then } 0 = \frac{1}{\sqrt{n}} \dot{l}_n^{(\epsilon)}(\hat{\beta}^{(\epsilon)}) = \frac{1}{\sqrt{n}} \dot{l}_n^{(\epsilon)}(\beta_0) - \left(-\frac{\ddot{l}_n^{(\epsilon)}(\beta_n^*)}{n}\right) \sqrt{n}(\hat{\beta}^{(\epsilon)} - \beta_0), \text{ where } |\beta_n^* - \beta_0| \leq |\hat{\beta}^{(\epsilon)} - \beta_0|.$$

$$\text{Note that } \ddot{l}^{(\epsilon)}(\beta|X_i) = \epsilon_i \ddot{l}(\beta|X_i), \text{ thus } -E_0 \left\{ \ddot{l}_{jk}^{(\epsilon)}(\beta_0|X) \right\} = -E_0 \left\{ \ddot{l}_{jk}(\beta_0|X) \right\} = I(\beta_0), \\ -\frac{1}{n} \ddot{l}_n^{(\epsilon)}(\beta_n^*) = -\frac{1}{n} \ddot{l}_n^{(\epsilon)}(\beta_0) + o_p(1) \xrightarrow{p} I(\beta_0).$$

Therefore,

$$\begin{aligned} \sqrt{n}(\hat{\beta}^{(\epsilon)} - \beta_0) &= I^{-1}(\beta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_n^{(\epsilon)}(\beta_0|X_i) + o_p(1) \\ &= I^{-1}(\beta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_n(\beta_0|X_i) \cdot \epsilon_i + o_p(1). \end{aligned}$$

Take the specific case of a Bernoulli distribution. If  $X_i$ 's are i.i.d. Bernoulli( $p$ ), then the MLE  $\hat{p} = \frac{\sum_{i=1}^n X_i}{n}$ . The perturbed version of  $\hat{p}$  takes the form  $\hat{p}^{(\epsilon)} = \frac{\sum_{i=1}^n X_i \epsilon_i}{\sum_{i=1}^n \epsilon_i}$ .

Notice that  $a_1$  represents an adjustment due to estimating  $F_{k_0}^*(s) = Pr(S \leq s | W = w_k, \delta = 1) = p_{k_0}$  by  $F_{N_k} = \frac{\sum_i^N \delta_i I(W_i = w_k) I(S_i \leq s)}{\sum_i^N \delta_i I(W_i = w_k)}$ . Conditional on  $W = w_k$  and  $\delta = 1$ ,

$I(S_i \leq s) \sim \text{Bernoulli}(p_{k_0})$ . Thus following the above arguments:

$$\begin{aligned} \text{if } \sqrt{N} [F_{N_k}(s) - F_{K_0}^*(s)] &= \frac{1}{N} \sum_{i=1}^N \phi_{F_{k_0}^*}(s)(\delta_i, S_i, W_i) + o_p(1), \\ \text{then } \sqrt{N} [F_{N_k}^{(\epsilon)}(s) - F_{K_0}^*(s)] &= \frac{1}{N} \sum_{i=1}^N \phi_{F_{k_0}^*}(s)(\delta_i, S_i, W_i) \cdot \epsilon_i + o_p(1), \text{ where } F_{N_k}^{(\epsilon)} = \\ &= \frac{\sum_i^N \delta_i I(W_i = w_k) I(S_i \leq s) \epsilon_i}{\sum_i^N \delta_i I(W_i = w_k) \epsilon_i}. \end{aligned}$$

Following similar arguments for  $\dot{\Psi}_\beta^{(\epsilon)} = \dot{\Psi}_\beta$ , it is straightforward to show  $\dot{\Psi}_{F_k^*}^{(\epsilon)}(\beta_0, F_0^*) = \dot{\Psi}_{F_k^*}(\beta_0, F_0^*)$ .

Thus

$$\begin{aligned} a_1^{(\epsilon)}(\delta_i, S_i, W_i) &= \sum_1^k \dot{\Psi}_{F_k^*}^{(\epsilon)}(\beta_0, F_0^*) [F_{Nk}^{(\epsilon)}(s) - F_{K_0}^*(s)] \\ &= \sum_1^k \dot{\Psi}_{F_k^*}(\beta_0, F_0^*) [F_{Nk}(s) - F_{K_0}^*(s)] \cdot \epsilon_i + o_p(1) \\ &= a_1(\delta_i, S_i, W_i) \cdot \epsilon_i + o_p(1). \end{aligned}$$

#### 4. To show that $a_2^{(\epsilon)} = \epsilon a_2 + o_p(1)$

Let  $\pi(Y, Z, W, \alpha)$  denote  $P(\delta = 1|Y, Z, W)$ . We estimate  $\alpha$  by maximizing  $\text{Log}L(\alpha|Y, Z, W) = \sum_{i=1}^n \delta_i \log \pi(Y_i, Z_i, W_i; \alpha) + (1 - \delta_i) \log \{1 - \pi(Y_i, Z_i, W_i; \alpha)\}$ . Therefore  $\hat{\alpha}$  is a MLE estimator and following similar arguments in 3., we get that  $\sqrt{N} \dot{\Psi}_\alpha(\beta_0, F_0^*, \pi(\alpha_0))[\hat{\alpha} - \alpha_0] = \frac{1}{\sqrt{N}} \sum_{i=1}^N a_2(\delta_i, Y_i, Z_i, W_i) + o_p(1)$ ,  $\sqrt{N} \dot{\Psi}_\alpha^{(\epsilon)}(\beta_0, F_0^*, \pi(\alpha_0))[\hat{\alpha}^{(\epsilon)} - \alpha_0] = \frac{1}{\sqrt{N}} \sum_{i=1}^N a_2^{(\epsilon)}(\delta_i, Y_i, Z_i, W_i) + o_p(1)$ , and  $a_2^{(\epsilon)}(\delta_i, Y_i, Z_i, W_i) = a_2(\delta_i, Y_i, Z_i, W_i) \cdot \epsilon_i + o_p(1)$ .

Therefore, the proof for (10) is completed. Furthermore,

$$\sqrt{n}(\hat{\beta}^{(\epsilon)} - \hat{\beta}) = -[\dot{\Psi}_\beta]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_1(\delta_i, Y_i, S_i, Z_i, W_i)(\epsilon_i - 1) + o_p(1).$$

## C.2

In the this section, we show that

$$\sqrt{N}(\hat{\gamma}^{(\epsilon)}(\hat{\alpha}^{(\epsilon)}) - \hat{\gamma}(\hat{\alpha})) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_2(\delta_i, Y_i, S_i(1), Z_i, W_i) \cdot (\epsilon_i - 1) + o_p(1).$$

Similar to the arguments used in Section C.1,

$$\sqrt{N} \begin{pmatrix} \hat{\gamma}^{(\epsilon)}(\alpha_0) - \gamma_0 \\ \hat{\alpha}^{(\epsilon)} - \alpha_0 \end{pmatrix} = \sqrt{N} \begin{pmatrix} \frac{1}{N} \sum_{i=1}^N \frac{\delta_i}{\pi_{0i}} \tilde{l}_0(S_i) \cdot \epsilon_i \\ \frac{1}{N} \sum_{i=1}^N \tilde{l}_0^\alpha \cdot \epsilon_i \end{pmatrix}. \quad (11)$$

Moreover

$$\frac{1}{N} \sum_{i=1}^N \left( \frac{\delta_i}{\hat{\pi}_i} - \frac{\delta_i}{\pi_0} \right) \tilde{l}_0(S_i) \cdot \epsilon_i = o_p(1) - \frac{1}{N} \sum_{i=1}^N \delta_i \tilde{l}_0(S_i) \cdot \epsilon_i \left[ \frac{\pi_0^T}{\pi^2} (\hat{\alpha} - \alpha_0) \right].$$

Then

$$\sqrt{N} [\hat{\gamma}^{(\epsilon)}(\hat{\alpha}^{(\epsilon)}) - \gamma_0] = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_2(\delta_i, Y_i, S_i(1), Z_i, W_i) \cdot \epsilon_i + o_p(1).$$

Therefore

$$\sqrt{N}(\hat{\gamma}^{(\epsilon)}(\hat{\alpha}^{(\epsilon)}) - \hat{\gamma}(\hat{\alpha})) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_2(\delta_i, Y_i, S_i(1), Z_i, W_i) \cdot (\epsilon_i - 1) + o_p(1).$$

Conditional on the data and given that  $E[\epsilon] = 1$ ,  $Var[\epsilon] = 1$ ,  $E[(\epsilon - 1)^2] = 1$ , we have  $E[\phi_0(\epsilon - 1)] = 0$ ,  $Var[\phi_0(\epsilon - 1)] = Var[\phi_0]$ ,

$$cov(\phi_1(\delta, Y, S, Z, W)(\epsilon - 1), \phi_2(\delta, Y, S, Z, W)(\epsilon - 1)) = cov(\phi_1(\delta, Y, S, Z, W), \phi_2(\delta, Y, S, Z, W)).$$

Therefore, conditional on the data

$$\sqrt{n} \begin{pmatrix} \hat{\beta}^{(\epsilon)} \\ \hat{\gamma}^{(\epsilon)} \end{pmatrix} - \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \xrightarrow{d} N \left( \mathbf{0}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right). \quad (12)$$

Because  $\log \hat{R}R^{(\epsilon)}$  is a continuous function of  $\hat{\beta}^{(\epsilon)}$  and  $\hat{\gamma}^{(\epsilon)}$ , the regular delta method applies and the unconditional distribution of  $\sqrt{N} \{ \log \hat{R}R - \log RR_0 \}$  can be approximated by  $\sqrt{N} \{ \log \hat{R}R^{(\epsilon)} - \log \hat{R}R \}$  conditional on the observed data.

## D Simulation results from the BIP-only design

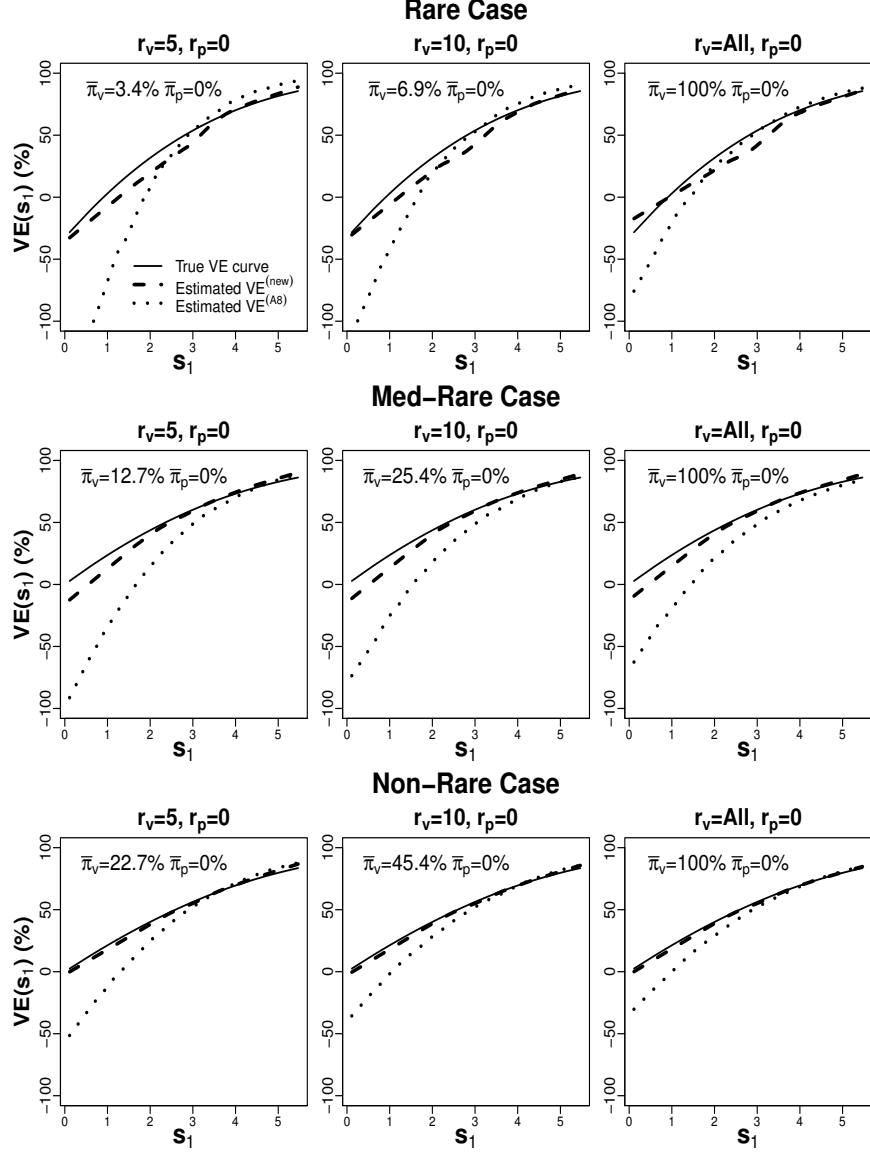


Figure 1: Estimated vaccine efficacy curve using our proposed method without assumption A8 and using the HGW method with assumption A8, compared to the true VE curve for checking the bias of these two estimators based on 500 simulated datasets for the Rare case where the probability of  $Y = 1$  for  $Z = 0$  ( $r_0$ ) equals 0.090 and for  $Z = 1$  ( $r_1$ ) equals 0.042, the Med-Rare case where  $r_0 = 0.055$  and  $r_1 = 0.020$ , and the Non-Rare case where  $r_0 = 0.0090$  and  $r_1 = 0.0068$  with a BIP-only design.

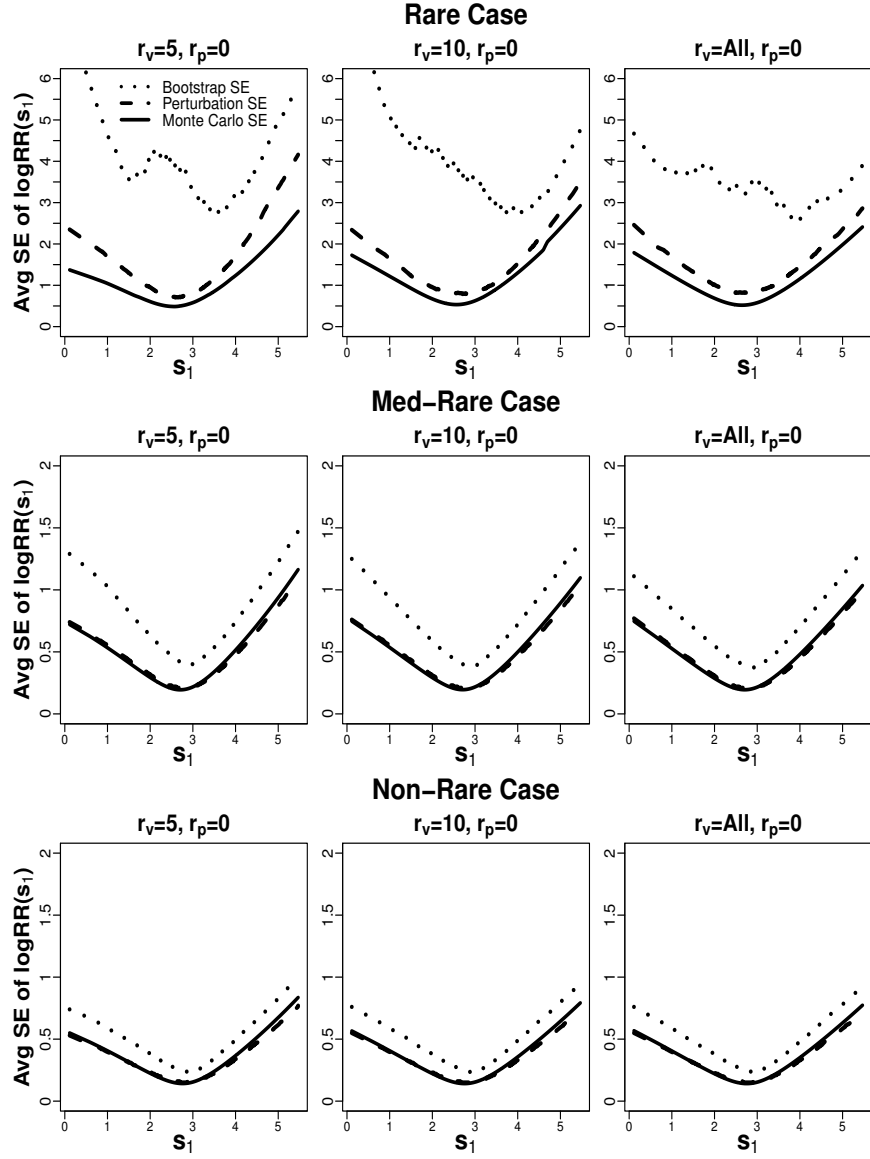


Figure 2: Estimated standard errors of  $\log\widehat{RR}(s_1)$ , solid for the Monte Carlo SEs, dashed for the perturbation resampling approach and dotted for the bootstrap approach, for the Rare case, the Med-Rare case, and the Non-Rare case with a BIP-only design.

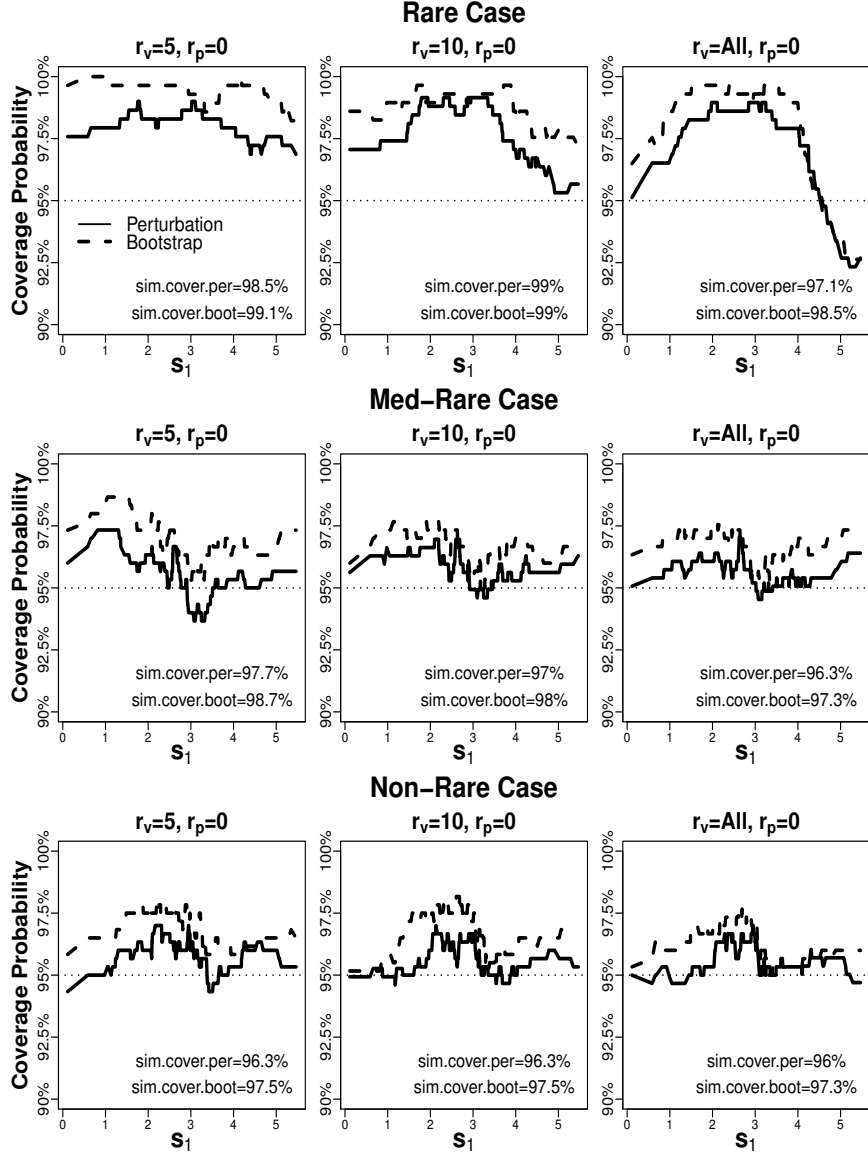


Figure 3: Empirical coverage probabilities of 95% pointwise confidence intervals and simultaneous confidence bands about  $VE(s_1)$ , for the Rare case, the Med-Rare case, and the Non-Rare case with a BIP-only design.

The relative efficiency, defined as ratio of sampling variance of  $\hat{\beta}^{(new)}$  and sampling variance of  $\hat{\beta}^{(A8)}$ , was 1.23, 1.11, 1.32, and 1.12 for the Med-Rare Case BIP-only design with  $r_v = 10$  and  $r_p = 0$  for  $\beta_0, \beta_1, \beta_2, \beta_3$ , respectively.